

---

## PROXIMATE SOURCES OF POPULATION SEX IMBALANCE IN INDIA\*

EMILY OSTER

*There is a population sex imbalance in India. Despite a consensus that this imbalance is due to excess female mortality, the specific source of this excess mortality remains poorly understood. I use microdata on child survival in India to analyze the proximate sources of the sex imbalance. I address two questions: when in life does the sex imbalance arise, and what health or nutritional investments are specifically responsible for its appearance? I present a new methodology that uses microdata on child survival. This methodology explicitly takes into account both the possibility of naturally occurring sex differences in survival and possible differences between investments in their importance for survival. Consistent with existing literature, I find significant excess female mortality in childhood, particularly between the ages of 1 and 5, and argue that the sex imbalance that exists by age 5 is large enough to explain virtually the entire imbalance in the population. Within this age group, sex differences in vaccinations explain between 20% and 30% of excess female mortality, malnutrition explains an additional 20%, and differences in treatment for illness play a smaller role. Together, these investments account for approximately 50% of the sex imbalance in mortality in India.*

India has a serious population sex imbalance. There are around 108 men for every 100 women in the country as a whole. In a country with the same level of development and typical mortality patterns, one would expect to see about 100 men for every 100 women. Sen (1990, 1992) coined the phrase “missing women” to describe this population imbalance, and attributed it to sex discrimination. Consistent with this view, other authors (Kishor 1993; Visaria 1971) have argued, based on census data and other sources, that the sex imbalance is almost certainly due to excess female mortality.

There is a very large literature on the underlying sources of parental sex preferences (see, e.g., Agnihotri 2000; Agnihotri, Palmer-Jones, and Parikh 2002; Murthi, Mamta, and Dreze 1995; Qian 2008; Rosenzweig and Schultz 1982) that focuses on the relative contributions of factors such as female labor-force participation and female education in determining overall sex ratios. A second literature, more closely related to this work, focuses on the proximate sources of female mortality<sup>1</sup>: that is, conditional on preferences, what specific treatments (or lack thereof) are responsible for the differences in mortality (Basu 1989; Borooah 2004; Griffiths, Matthews, and Hinde 2002; Mishra, Roy, and Retherford 2004; Pande 2003).

Despite this second literature, a coherent overall picture of the proximate sources of excess female mortality is still lacking. This article focuses on two primary questions: at what ages does most of the excess female mortality occur, and what is the relative contribution of various forms of neglect to this excess mortality? In contrast to most of the existing

---

\*Emily Oster, University of Chicago and NBER. Direct correspondence to Emily Oster, University of Chicago, Department of Economics, 1126 E. 59th Street, Chicago, IL 60637; e-mail: eoster@uchicago.edu. Gary Becker, Kerwin Charles, Steve Cicala, Amy Finkelstein, Andrew Francis, Jon Guryan, Matthew Gentzkow, Lawrence Katz, Michael Kremer, Steven Levitt, Kevin Murphy, Jesse Shapiro, Andrei Shleifer, Rebecca Thornton, and participants in seminars at Harvard University, the University of Chicago, and NBER provided helpful comments. I am grateful for funding from the Belfer Center, Kennedy School of Government. Laura Cervantes provided outstanding research assistance.

1. Throughout the article, I refer to “proximate sources” of excess female mortality. I define a proximate source as an investment that differs across sexes. For example, if vaccination levels are higher for boys than for girls, vaccination is likely to be one proximate source of excess mortality.

literature, I am concerned not only with whether various health and nutrition inputs play a role, but also with how large that role is.

The methodology used here, formally outlined in the following section, differs from most of the previous literature in two ways. First, I use data from Africa on sex differences in mortality and child health investments as a comparison for India. Existing literature (e.g., Das Gupta 1987) has often focused solely on sex differences in mortality in India. However, because when boys and girls receive equal treatment by their parents or caregivers, boys are more likely to die, the lack of a comparison group likely understates the extent of excess female mortality. Second, when considering the proximate sources of excess female mortality in childhood, I consider not only the difference in treatment but also the importance of that treatment for mortality (i.e., the difference in mortality probability with and without treatment). Multiplying these two factors gives full information about the importance of each element for understanding the overall excess female mortality. The literature generally has considered only the difference across sexes in each treatment; it has not considered the importance of these treatments in mortality, which is crucial for evaluating the relative contribution of each input (Basu 1989; Borooah 2004; Griffiths et al. 2002; Mishra et al. 2004; Pande 2003).<sup>2</sup>

I first use microdata to identify exactly the age source of the excess female mortality in childhood and to explore the importance of childhood sex bias in the overall imbalance. This question has, of course, been addressed by other researchers (Das Gupta and Bhat 1997; Dyson 1984; Klasen 1994; Padmanabha 1982; Preston and Bhat 1984); the work here uses a new methodology, but the results largely echo what has been found in the previous literature. In particular, the results suggest important variations within young children. All areas of India see relatively little excess female mortality between the ages of a few months and 2 years, yet substantial excess mortality between 2 and 5 years of age. I also present evidence on the contribution of the under-5 sex ratio bias to the overall bias. Using demographers' life tables (Coale, Demeny, and Vaughn 1983), I calculate the expected sex ratio overall in India, assuming the empirically observed sex ratio at 5 years of age, and normal mortality thereafter. This exercise suggests that virtually all the sex ratio imbalance in the country can be explained by excess under-5 mortality.

Following this analysis, I move on to the primary contribution of the article, exploring the proximate sources of this excess female mortality between the ages of 2 and 5. Consistent with previous literature, I focus on biases in nutrition, preventative medicine, and medical treatment. The evidence here suggests that, contrary to some of the previous literature, sex differences in vaccinations play a very large role in the sex imbalance, explaining about 20% to 30%. Malnutrition explains about 20%. Interestingly, differences in treatment for respiratory infections and diarrhea together explain only about 5% of the imbalance, and approximately 50% is left unexplained by these childhood investments.

The results here have potentially important policy implications, suggesting that increases in vaccinations for girls could have a large effect on the overall sex imbalance in India.

## METHODOLOGY

Here, I discuss the methodology used for estimating both the overall sex imbalance in mortality by age and the contribution of various investments to this sex imbalance. To illustrate the basic concept, define  $D$  as the differences between sexes in some investment (e.g., the difference in the chance of receiving a measles vaccination). I define  $\mu$  as the importance of this investment in mortality (e.g., the difference in mortality probability if vaccinated

---

2. This should not necessarily be taken as a criticism of this work. Generally, the authors did not intend to calibrate the importance of different explanations in the sex bias, but rather to demonstrate that one particular explanation might play a role.

and unvaccinated) and  $\psi$  as the overall excess female mortality. The share of the overall difference that is explained by this investment is, therefore

$$\frac{D\mu}{\Psi}. \quad (1)$$

That is, the overall contribution is simply the expected excess mortality resulting from differences in this particular investment ( $D\mu$ ) divided by the total excess mortality. The challenge, then, is estimating  $D$ ,  $\mu$ , and  $\Psi$ .

First, consider the estimation of overall differences in mortality:  $\Psi$  from Eq. (1). This is an input to understanding the importance of different investments but is also—when calculated for each age group—the parameter that indicates the importance of each age group in the overall excess female mortality. This variable is, intuitively, the difference between actual and expected probability of death for girls. In other words,  $\Psi$  measures how much more likely a girl is to die relative to what is expected, based on mortality of boys.<sup>3</sup> Perhaps the most obvious way to estimate this would be to simply calculate the difference between male and female mortality in India and assume that  $\Psi$  is equal to that difference; in other words, assume that girls have the same mortality as boys. The problem with this method, however, is that differences may well exist between sexes even in nondiscriminatory environments. If these differences exist, simply comparing the two sexes within India may understate (or overstate) the excess female mortality.<sup>4</sup> To solve this problem, I employ a “difference-in-difference” technique, in which I use data on India and a comparison region (sub-Saharan Africa) to evaluate the difference in mortality in India relative to the “expected” difference. The equation estimated is

$$\text{Pr}(\text{die}) = (\alpha) + (\beta_0)(\text{girl}) + (\beta_1)(\text{India}) + (\beta_2)(\text{girl} \times \text{India}) + \Phi\mathbf{X}. \quad (2)$$

This regression relates the probability of death to child characteristics.  $\Phi\mathbf{X}$  is simply a vector of controls—for example, mother’s education, family income, and other variables—that may affect child mortality.  $\alpha$  is a constant in the regression. If  $\mathbf{X} = 0$  (i.e., the value of all controls is equal to zero), then  $\alpha$  is equal to the probability of death for a boy in the comparison region. The  $\beta$  coefficients measure differences in the probability of death across sex and area:  $\beta_0$  is the difference in probability of death for girls versus boys in the comparison region, and  $\beta_1$  is the difference in probability of death (on average) between India and the comparison region.

The coefficient of interest is  $\beta_2$ , the interaction between being a girl and living in India. The coefficient on this interaction is the sex imbalance in mortality. If India is similar to the comparison region, then I should find  $\beta_2 = 0$ . If girls are disadvantaged, I should find  $\beta_2 > 0$ . In the language of Eq. (1),  $\beta_2 = \Psi$ .<sup>5</sup>

Eq. (2) illustrates the problem with comparing death rates for boys and girls within India and using that comparison to measure excess female mortality. If I estimate Eq. (2) using only data from India, then India = 1 for all observations. In that case, I will not be

3. The concept of an expected death rate based on male mortality presumes some standard mortality schedules to which the actual mortality can be compared. In practice, it is not possible to perfectly identify the biological expected mortality paths for men and women. Empirically, I rely on data from a less-discriminatory environment to provide information on the “expected” relative mortality for boys and girls. This is discussed more extensively in the data section.

4. It is very frequently observed that men are more likely to die at all ages in nondiscriminatory environments, but the reasons are not obvious. Wells (2000) provides good links to the literature on the existence of this effect and presents one potential explanation.

5. Note that although I continue to refer to this as a difference-in-difference regression, this term does not connote anything about identification. This regression is simply a mechanical adjustment for baseline differences between the sexes.

able to separately identify  $\beta_2$  and  $\beta_0$ , and I will observe that the coefficient on *girl*—the  $\Psi$  I am interested in—is equal to  $\beta_2 + \beta_0$ . That is, I will not be able to separate the effect of sex overall from the effect of sex in India, and the coefficient measuring excess female mortality will not be interpretable as such.

The second methodological issue is identification of  $D\mu$ . I focus on two primary analyses: individual regression (which will estimate the entire quantity  $D\mu$ ) and direct calculation of  $D$  and  $\mu$  separately. Consider first the individual regression. Imagine that I have an individual-level panel in which I observe the level of health investment and mortality outcomes for children. I can then estimate the quantity  $D\mu$  by comparing the coefficient on *girl*  $\times$  *India* in two difference-in-difference regressions: the first without controls for the health investment, and the second with these controls. In particular, denoting the health investment as  $Z$ , I first estimate Eq. (2) and then Eq. (3):

$$\text{Pr}(\text{die}) = \delta + \gamma_0(\text{girl}) + \gamma_1(\text{India}) + \gamma_2(\text{girl} \times \text{India}) + \gamma_3(Z) + \Phi\mathbf{X}. \quad (3)$$

Given these regressions,  $D\mu = \beta_2 - \gamma_2$ .

Perhaps the easiest way to see the intuition behind this calculation is to think of  $Z$  as an omitted variable in Eq. (2).  $\beta_2$  captures the effect of many investments, one of which is  $Z$ . By not controlling directly for  $Z$ ,  $\beta_2$  is upward biased. Controlling for  $Z$  will decrease  $\beta_2$ , with the amount depending on how important  $Z$  is in explaining the mortality imbalance. More concretely, I model the relationship between  $Z$  and the interaction by using Eq. (4):

$$Z = \eta + v_1(\text{girl}) + v_2(\text{India}) + v_3(\text{girl} \times \text{India}). \quad (4)$$

Based on the omitted variable intuition,  $\beta_2 = \gamma_2 + (\gamma_3)(v_3)$ . This means that  $\beta_2 - \gamma_2 = (\gamma_3)(v_3)$ . From this, it is straightforward to see why  $(\gamma_3)(v_3)$  is an estimate of  $D\mu$ :  $\gamma_3$  is just a measure of the effect of health investment on mortality ( $\mu$ ), and  $v_3$  is a measure of the sex bias in that investment ( $D$ ). The product of these two gives the share explained by that particular investment. As noted, this analysis requires an individual-level panel data set (or enough information to construct one).

A significant concern with this approach is that the elements of  $Z$  may overcontrol and thus soak up some of the effect of parental preferences. If  $Z$  measures vaccination, but differences in vaccination are simply a proxy for preferences and are perfectly correlated with all other forms of discrimination, then the difference between  $\beta_2$  and  $\gamma_2$  will capture much more than just the effect of vaccination. However, this will be an issue only if vaccination overall is correlated with parental sex preferences *and* if mortality is correlated with sex preferences, which does not seem to be the case empirically.<sup>6</sup> Regressing mortality and vaccination on the parental reported ideal sex ratio (parents are asked about their ideal number of sons and daughters in the later survey waves) yields insignificant and small coefficients (results available from the author). Despite this, a potential concern is omitted parental preferences. One way to partially adjust for possible preference differences is to include some simple preference controls—in particular, the mother's reported ideal sex ratio. When I do this, the results do not change. Of course, this control may not fully capture preferences, and this omission remains a concern. One advantage of the second methodology (discussed later in the article) is that these concerns will be largely avoided.

6. To see this, consider the discussion of the omitted variable bias intuition in the preceding text. To say that one "overcontrols" implies that the adjustment between the two regressions is too large: that is,  $(\gamma_3)(v_3)$  is too big. This will be the case if  $\hat{\gamma}_3$ —the relationship between vaccination and mortality—is overestimated. Based on the standard arguments concerning omitted variable bias, omitting parental preferences will be a problem if measles vaccination is correlated with preferences and mortality is correlated with preferences.

The second methodology used to calculate  $D\mu$  is direct estimation of  $D$  and  $\mu$ . I first use the National Family and Health Survey (NFHS) to directly calculate the sex differences in treatment by estimating Eq. (4): the estimate of  $D$  is simply  $v_3$ . I then obtain estimates of  $\mu$  from the existing literature, based on studies in which mortality outcomes were observed for children with varying levels of health investment. There are two advantages to this approach. First, because the estimates of the effect of treatment on mortality come from other surveys, there is less concern about bias arising from this particular sample. Second, this technique allows me to obtain estimates for the effect of nutrition and medical treatment, as well as for vaccination. As a final robustness check, I replicate the individual-level analysis by using data at the regional level. Although this methodology is likely to be the least appealing, due to the other differences I expect to see across regions, it does allow me to simultaneously control for all of the elements of mortality.

Before moving on to the data and results, it is worth briefly discussing how the methodology used here differs from that of previous studies. There are two basic differences. First, most studies (e.g., Das Gupta 1987) have used only a difference approach—comparing the death rates of boys and girls in India. This will generally underestimate true excess female mortality because boys are more likely to die in nondiscriminatory environments. Second, studies on proximate sources of excess mortality generally focus on estimating the differences in treatment by sex (i.e.,  $D$  from the earlier discussion) and not the effect of these treatments on mortality (Basu 1989; Borooah 2004; Griffiths et al. 2002; Mishra et al. 2004; Pande 2003). Without adjusting for differences in  $\mu$ , it is very difficult to say anything conclusive about which inputs are more important in explaining the sex differences.<sup>7</sup>

## DATA

The analyses here are run using individual-level microdata on child survival and health investments.<sup>8</sup> For India, the data used are from two waves of the NFHS (1992–1993 and 1998–1999), which covers approximately 90,000 women in each wave. Women are asked about their birth history, including children ever born; dates of birth; whether every child is alive; and if not, when the child died. In addition, for children younger than age 5, information is collected on vaccination, medical treatment, and malnutrition. In the 1992–1993 survey, vaccination information was collected for all children, including those who had died. In the 1998–1999 survey, however, this information was not collected for children who had died. The analyses of vaccination, therefore, include only the 1992–1993 round of the NFHS.

As discussed in the methodology section, the size of the sex imbalance in mortality and investments in India is evaluated relative to the size of this imbalance in a comparison area. This allows me to difference out any differences across sexes (favoring boys or girls) that occur in apparently nondiscriminatory (or less-discriminatory) environments. The literature on the “missing women” suggests two natural comparisons: sub-Saharan African (Sen 1990, 1992) and demographer’s life tables (Coale 1991; Klasen and Wink 2002). Sex differences in mortality in sub-Saharan Africa are similar to those predicated in the life tables, suggesting that these two comparisons will give similar results. The advantage of using sub-Saharan Africa, as I do here, is that the same type of microdata on children is available from a number of countries. The Demographic and Health Surveys

7. In the existing literature, there is much focus on the differences between North and South India. If I separate India into the two regions, I find that sex imbalances are higher in North India in virtually all of the inputs and in excess mortality, but the conclusions about patterns by age in childhood and about which proximate sources are most important holds.

8. This is in contrast to much of the literature on this topic, which relies on district-level data on sex ratios. The advantage of using the individual-level data is that I observe directly the relationship between health investments and mortality.



(DHS) in Africa mirror the NFHS, so the difference-in-difference analysis can be run at the level of the individual child. The comparison countries are Ethiopia, Kenya, Malawi, Namibia, Tanzania, and Zambia.

The three child investments analyzed are vaccination, malnutrition, and treatment for disease. There are seven possible vaccinations (three diphtheria, pertussus, tetanus [DPT] vaccines, two polio vaccines, a measles vaccine, and a *Bacillus Calmette-Guérin* [BCG] vaccine). In general, I use two measures of vaccination: the total number of vaccinations reported by the mother, and the total number marked on the child's health card. The results are extremely similar if I include dummy variables for each vaccination.

Information on malnutrition is based on actual height and weight measurements. Living children younger than age 4 in each household are measured and weighed. Their percentile weight-for-age is reported; weight-for-height and height-for-age are also reported, and all are very closely linked. I define children as severely malnourished if their percentile weight-for-age is less than 60% of the reference median for their age and sex. I use this indicator rather than a continuous measure because research on the effect of malnutrition on mortality indicates that mortality is largely unaffected by malnutrition higher than 60% of the reference median, but increases sharply below that (Chen, Chowdhury, and Huffman 1980).

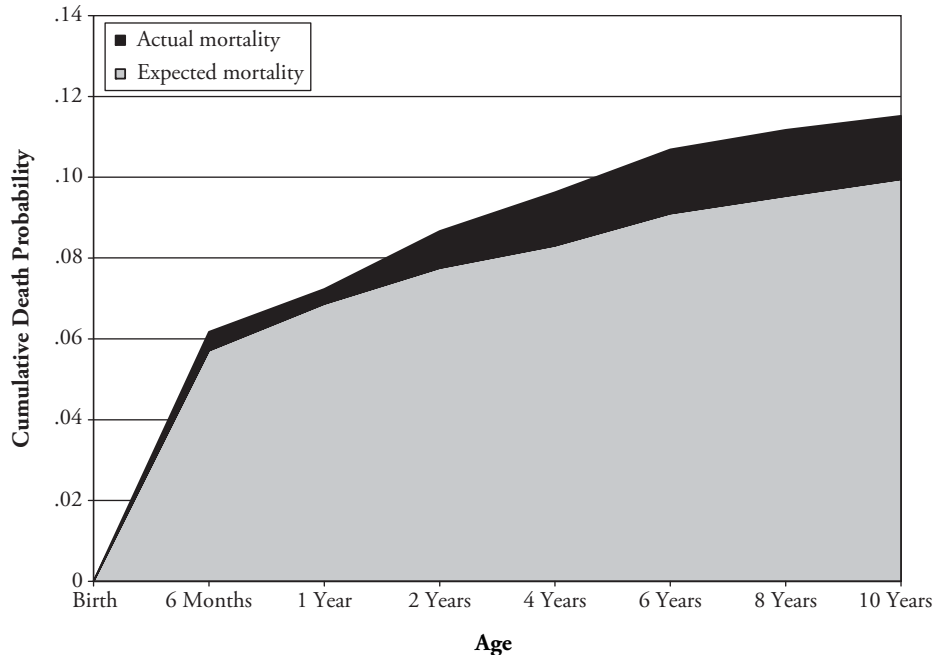
To evaluate differences in medical treatment, parents were asked whether each of their (living) children younger than 4 had diarrhea or symptoms of a respiratory infection in the past two weeks. If the answer was yes, they were asked what treatment was provided. I report children as having been treated if their parents reported having given the child any treatment (including a doctor visit and home remedies). In these data, differentiating by treatment type has little effect on the sex difference.

### AGE-SPECIFIC ORIGINS OF EXCESS FEMALE MORTALITY

The first set of results estimate the baseline excess female mortality. I estimate Eq. (2) for age groups ranging from birth to 10 years. The dependent variable is a series of indicators for having died within a particular age group. For example, the first variable is a 0-1 dummy variable for whether a child born in the past 10 years died before the age of 6 months; the second variable is a 0-1 dummy variable for whether the child died between 6 months and 1 year, conditional on having lived to 6 months of age. The additional age groups are 1-2 years, 2-4 years, 4-6 years, 6-8 years, and 8-10 years.

The results of this analysis can be seen graphically in Figure 1, which plots actual and expected mortality for girls in India by age group, with the expected mortality based on male mortality in India and the sex difference in Africa. By age 10, the actual probability of female deaths is almost 12%, compared with an expected probability of slightly less than 10%. Nearly all of this imbalance seems to arise between the ages of 1 and 4, when expected mortality is around 1.4% and actual mortality is a full 2.4%. The regression analog to this figure appears in Table 1, where the difference-in-difference estimate is the coefficient on *girl*  $\times$  *India*. Consistent with the picture, the difference is statistically significant between 6 months and 6 years, but not in the youngest or oldest groups.

Controls in this regression include child age, maternal age, maternal education, birth order dummy variables, and total number of siblings. In general, these enter with the expected sign and are unremarkable. However, the control for total number of siblings is worth a brief discussion. It is frequently suggested that one of the major reasons why female mortality is higher is that girls are, on average, in larger families (because of some form of sex-biased fertility stopping rule). This seems to be somewhat true because excluding the control for number of siblings leads to a larger estimate for the interaction between sex and India. In this sense, I can say that one proximate source of excess mortality is larger family size. However, exactly what health investments are denied in larger families remains to be analyzed.

**Figure 1. Actual and Expected Female Mortality in India**

*Notes:* The sample is limited to children born in the 10 years before the survey, and estimates are based on the regressions shown in Table 1. Expected mortality is the predicted mortality from the regression for girls in India if they did not have the negative effect of being a girl in India (i.e., if the value of the interaction term was zero). Actual mortality is actual predicted mortality.

The results in Table 1 give a sense of the magnitude of excess female mortality in childhood and the periods of childhood that are most crucial. A related question is how important childhood is in explaining the overall sex imbalance. To get a sense of this issue, I calculate the predicted sex ratio in the population (based on life tables), taking the sex ratio at age 5 as given. If the predicted sex ratio in the population based on this calculation is much lower than the actual sex ratio, then it suggests that any excess mortality up to age 5 is probably unimportant in the overall sex bias. In contrast, if the predicted and actual sex ratios are similar, it would suggest that excess female mortality before age 5 explains most of the overall sex imbalance.

The result of these calculations appears in Table 2 (details of the calculation are available in Appendix A, on *Demography's* Web site), and the results suggest that a very large share of the sex bias can be explained by events occurring up to age 5. This, in turn, suggests that understanding the proximate sources of mortality in this age bracket may go far in helping understand the overall problem. This is not surprising. Mortality rates among young children are much higher than among prime-age adults, so one would expect mortality in childhood to contribute to a large share of the sex imbalance simply because the level is higher. It is worth noting, however, that the relationship is not mechanical. Even though mortality from 0 to 6 months is much higher than mortality later in childhood, that period does not contribute very much to the sex imbalance.

**Table 1. Sex Imbalance in Death, by Age, for All of India (dependent variable is child died in a given age range [0/1])**

Explanatory Variables	< 6 Months	6 Months– 1 Year	1–2 Years	2–4 Years	4–6 Years	6–8 Years	8–10 Years
Girl	–0.0104** (–6.67)	0.0002 (0.33)	–0.0011 <sup>†</sup> (–1.81)	–0.0014** (–2.66)	–0.0000 (–0.04)	–0.0000 (–0.02)	0.0007 (1.21)
Girl × India	0.0016 (0.91)	0.0021** (2.86)	0.0055** (6.48)	0.0041** (5.57)	0.0026** (3.23)	0.0005 (0.70)	–0.0006 (–0.91)
India	–0.0077** (–5.00)	–0.0135** (–16.12)	–0.0175** (–18.61)	–0.0098** (–12.65)	–0.0088** (–10.29)	–0.0032** (–4.58)	–0.0008 (–1.16)
Wealth (durables)	–0.0049** (12.77)	–0.0012** (–6.96)	–0.0019** (–9.21)	–0.0020** (–11.14)	–0.0029** (–13.98)	–0.0015** (–8.19)	–0.0012** (–6.63)
Child's Age	–0.0002** (–6.61)	–0.0004** (–3.89)	–0.0000 (0.42)	–0.0000 (–0.39)	0.0007** (5.56)	0.0003** (2.12)	0.0010** (4.64)
Mother's Age	–0.0026** (–27.69)	–0.0004** (–7.42)	–0.0005** (–8.90)	–0.0002** (–4.95)	–0.0003** (–6.28)	–0.0001 (–1.31)	–0.0000 (–0.46)
Mother's Education	–0.0014** (–13.27)	–0.0004** (–7.46)	–0.0007** (–10.85)	–0.0003** (–6.43)	–0.0003** (–6.33)	–0.0002** (–4.61)	–0.0000 (–0.24)
Number of Siblings	0.0257** (67.13)	0.0039** (19.33)	0.0045** (20.05)	0.0028** (15.42)	0.0025** (12.69)	0.0008** (5.26)	0.0005** (3.53)
Year	–0.0007** (–4.57)	–0.0004** (–5.97)	–0.0001 (–1.03)	0.0001 (1.50)	–0.0001 (–1.18)	0.0000 (0.16)	0.0000 (–0.42)
Number of Observations	338,943	294,410	278,397	243,496	211,896	150,801	86,604

Notes: All regressions are limited to children born in the 10 years before the survey. The dependent variable is a dummy variable equal to 1 if the child died within the specified age range, conditional on having lived until that age category. The measure of wealth is a measure of the number of durable goods owned; the durable goods are radio, television, refrigerator, motorcycle, bicycle, and car. Controls for birth order (dummy variables) are also included in all regressions. India is observed in 1992 and 1998. The African countries included are Ethiopia (2000), Kenya (1998, 2003), Malawi (2000), Namibia (1992, 2000), Tanzania (1996, 1999), and Zambia (1996, 2001). Numbers in parentheses are *t* statistics.

<sup>†</sup>*p* < .10; \**p* < .05; \*\**p* < .01

## PROXIMATE CAUSES OF EXCESS FEMALE MORTALITY

I turn now to estimating the importance of different health investments in explaining this excess female mortality in childhood. I separate this discussion into three parts, focusing first on the individual-level regression methodology, second on the direct calculation of medical effects, and third on an analysis at the regional level.

### Individual-Level Regression

The individual-level regression methodology will be useable only when considering the effect of vaccinations and only when using surveys from the first NFHS, in 1992. Information on malnutrition and medical treatment was not collected for children who have died, and later surveys did not ask about vaccinations for deceased children. Without this information, it is not possible to construct the necessary individual-level panel.

Table 3 compares the results of estimating Eqs. (2) and (3), which will evaluate the effect of vaccination differences on mortality differences. The regression is limited to children born between four and five years before the survey, and the dependent variable is a dummy variable for having died between 18 months and 4 years, conditional on having lived to 18 months. The sample size is smaller than for the similar age group in Table 1



**Table 2. Share of Overall Imbalance Explained by Age 5**

	All India
Observed Sex Ratio (F / M) at Age 5	0.914
Predicted Sex Ratio, Model Mortality After Age 5	0.933
Empirical Population Sex Ratio	0.932

*Notes:* Sex ratio is the number of women divided by the number of men. Observed sex ratio at age 5 is from the NFHS. The predicted population sex ratio is calculated by assuming that the sex ratio at age 5 is naturally occurring and then assuming that mortality is equal to what is reported in life tables after age 5. Mortality is taken from the Coale, Demeny, and Vaughn (1983) life tables, using the West model mortality level 17 with a gross reproductive ratio of 2.0. Empirical population sex ratio is from the 1991 census.

**Table 3. Impact of Vaccines on Excess Female Mortality (dependent variable is child died at age 1–4 years)**

Explanatory Variables	(1)		(2)	
	Coefficient	<i>t</i> Statistic	Coefficient	<i>t</i> Statistic
Girl	−0.0013	−0.44	−0.0007	−0.29
Girl × India	0.0091*	−2.33	0.0065*	2.03
India	−0.0233**	−5.58	−0.0393**	−8.13
Number of Vaccinations Reported by Mother			−0.0009**	−3.72
Number of Vaccinations on Health Card			−0.0026**	−6.57
Number of Observations	13,817		13,817	
Share Explained (%)			28.4	

*Notes:* Controls in all regressions include dummy variables for child's size at birth, mother's age, mother's education, child's age, income (durables), total number of children, and dummy variables for birth order. The two measures of vaccines capture both the total number of vaccines reported by the mother and the total number marked on the health card (three DPT vaccines, two polio vaccines, measles, and BCG). Data come from the 1992–1993 NFHS.

\* $p < .05$ ; \*\* $p < .01$

for two reasons: first, I use only children born four to five years prior to the survey, not all children born in the past 10 years; and second, I use only the data from the 1992 NFHS because the 1998 NFHS did not ask about vaccinations for children who were deceased. Column 1 estimates Eq. (2), and column 2 estimates Eq. (3). As discussed in the methodology section, the share explained is calculated as the difference in the interaction coefficient divided by the interaction coefficient in column 1. Vaccinations have a significant negative effect on mortality. Moving from zero vaccinations to complete vaccination decreases the probability of dying between ages 1 and 4 by a full 1.8%. In addition, vaccinations seem to explain a large share of the sex imbalance: approximately 30%. The standard errors are sufficiently large that I cannot reject equality of the coefficients (i.e., I cannot reject that the amount explained is equal to zero). However, the size of the point estimate is certainly economically significant.

One possible weakness of this analysis is recall bias. The regression includes controls for both vaccinations reported on the health card and vaccinations reported by the mother. If mothers in India are less likely to remember vaccinations for girls who died, relative to boys who died, a potential bias exists. This would have the effect of omitting a measure

of true vaccination status while also including a measure of reported vaccination status. If true vaccination status (controlling for reported vaccination status) is correlated with the *girl × India* interaction and with mortality, then the coefficient may be biased. This concern is ameliorated, at least somewhat, by the inclusion of both measures of vaccination. Marks on the health card are likely to be a much better measure of actual vaccination status than maternal reports. And the closer these measures get to the true vaccination status, the less of a concern the omitted variable bias is. Further, including only the control for vaccinations reported on the health card makes relatively little difference in the results.

It is also possible that the effect of vaccinations varies by sex. If the benefit of vaccination for girls is larger than the benefit for boys, then the share of the bias explained by vaccination may be understated. Although there is some evidence on sex differences in the nonspecific protective effect of vaccinations (Aaby et al. 2002), these do not seem to be consistent across vaccines. As a sensitivity analysis, I repeat the regressions in Table 3, allowing for the effect of vaccination to differ by sex. The results (available from the author upon request) are virtually identical.

### Direct Calculation of Medical Effects

The second methodology here relies on direct evidence on the effect of child investments on mortality. In contrast to the regression framework, this analysis will be possible for all investments considered: malnutrition, treatment for diarrhea, treatment for respiratory infections, and vaccinations. However, I consider only measles vaccination because this is the illness for which we have the best estimates of the effectiveness of vaccination. Obviously, the effect of measles alone will be an understatement of the total vaccine effect.

The calculations here require two elements: the difference in treatment by sex ( $D$ ), and the effect of the treatment on mortality ( $\mu$ ). The first element is estimated in Table 4, which shows the sex bias in an indicator for severe malnutrition (Panel A), treatment for diarrhea (Panel B), treatment for respiratory infections (Panel C), and measles vaccination (Panel D). Controls are listed in the notes of the table. The results indicate that boys in India are about 1 percentage point less likely to be malnourished and that this effect is significant. The results on medical treatment are mixed: boys are significantly more likely to be treated for respiratory infections, but not any more likely to be treated for diarrhea. The largest observed effects are for measles vaccination; boys are approximately 7 percentage points more likely to be vaccinated.

Information on the second element—the effect of treatment on mortality—is presented in Web Appendix B. The details of the calculations appear in the Web Appendix, but in general, I use one of two techniques. In the case of malnutrition, I take advantage of studies in which nourishment levels of children were observed. The children were followed over time, and mortality outcomes were reported. The difference in mortality by nutritional status provides an estimate of the effect of malnutrition. In the case of treatment and vaccination, the effect is the product of the probability of dying from the illness (either diarrhea, acute lower respiratory infections, or measles) and the protective effect of treatment. In the case of measles, the effect of vaccination on mortality is the chance of dying from measles in India during this period multiplied by the effect of measles vaccination on measles mortality. The studies suggest that the protective effect of being well nourished is the largest, although the effect of measles vaccination is much larger than treatment for illnesses.<sup>9</sup> The studies used here are based on information from the developing world, or from India directly, so they should capture the experience of South Asia reasonably accurately.

9. The larger effect of malnutrition does not seem to be an artifact of the difference in methodology. Using the DHS data from Africa, it is possible to get an estimate of the effect of measles vaccination on mortality, which effectively parallels the estimate of malnutrition. The result suggests about a 3-percentage-point decrease in death probability with measles vaccination, similar to what I show in Web Appendix B. Although I do not use this estimate, because the goal is to use estimates from outside these data, it does provide some comfort.

**Table 4. Sex Imbalance in Malnutrition and Treatment of Illness**

Explanatory Variables	Coefficient	<i>t</i> Statistic
A. Treatment for Malnutrition		
Girl	−0.0025	−1.17
Girl × India	0.0116**	4.39
India	0.0342**	18.86
Number of observations	80,819	
B. Treatment for Diarrhea		
Girl	−0.0176	−1.46
Girl × India	−0.0027	−0.16
India	−0.1057**	−8.64
Number of observations	13,666	
C. Treatment for Respiratory Infections		
Girl	−0.00015	−0.21
Girl × India	−0.0357**	−3.24
India	0.0992**	12.55
Number of observations	26,077	
D. Measles Vaccination		
Girl	0.0356†	1.66
Girl × India	−0.0698**	−2.98
India	−0.4110**	−27.64
Number of observations	15,120	

*Notes:* Controls in all regressions include income, child's age, dummy variables for birth order, number of children in the household, mother's education, and mother's age. Data are from the 1992–1993 NFHS and the 1998–1999 NFHS. Panels A, B, and C include children aged 1–4 years; Panel D includes children aged 3–4 years.

†  $p < .10$ ; \*\*  $p < .01$

Table 5 brings together the results from Table 4 and Web Appendix B and presents them with reference to the size of the sex imbalance. The first row of the table shows the excess female mortality between ages 1 and 4 years. The share explained is simply the sex difference multiplied by the mortality effect, divided by this baseline difference. The results here suggest that food intake plays a sizable role in the sex imbalance (explaining 19%), but that treatment for diarrhea and respiratory infections plays only a limited role. The reason for this is straightforward. In the case of diarrhea, there is virtually no difference in treatment propensity. In the case of respiratory infections, there is a large difference in treatment propensity, but the chance of dying from that cause is not large and the protective effect of treatment is small. The effect of the measles vaccine provides a supportive robustness check on the earlier estimates of the effectiveness of vaccination from the individual-level regressions. Measles vaccination alone explains 21% of the sex imbalance. Although this is less than the 28% estimated in Table 3, it is an estimate for only one of many vaccinations.

### Regional-Level Analysis

The results in Tables 3 and 5 suggest that about 45% to 50% of the sex imbalance up to age 5 can be explained by vaccination, food intake, and medical treatment. One issue, however, is that these variables may not be independent. If malnutrition makes children more likely

**Table 5. Share of Missing Girls Explained by Food, Treatment, and Vaccination**

Variable	All India
Baseline Difference	0.0111
Treatment for Malnutrition (% explained)	19.1
Treatment for Diarrhea (% explained)	0.43
Treatment for Respiratory Infections (% explained)	3.9
Measles Vaccination (% explained)	20.6

*Notes:* The baseline difference is the value of the coefficient on the variable *girl*  $\times$  *India* in a regression of the form shown in Table 1, but aggregating death rates for children aged 6 months to 5 years. The share of the puzzle explained by each indicator is equal to the difference from Table 4, multiplied by the effect on mortality from Web Appendix B, and then divided by the baseline difference. In the case of respiratory infections and measles, the effect used from Web Appendix B is the average of the two study effects.

to die from measles, the effect of malnutrition in Web Appendix B is also partially an effect of measles vaccination. This may lead the results in Table 5 to overstate the total explanatory power of the investments considered. Without an individual-level panel in which one can observe all elements of food and treatment over time, this is a difficult problem to solve.

One option, however, is to collapse the data to the area level within India and then run regional-level equivalents to Eqs. (2) and (3). By doing the same analysis specified for the individual regression, I can infer the share of mortality explained by different investments. There are clear issues with this approach. States within India differ on many dimensions, and fully controlling for these differences may be difficult. (I attempt to do so with controls for education, durable good ownership and parental preferences, but fully controlling will be virtually impossible.) However, the advantage of the approach is that I can consider the effect of all investments simultaneously, which provides a useful robustness check.

The results of this analysis are in Table 6.<sup>10</sup> I show only the regression with no components of *Z* and the regression with all components of *Z*. What I will be able to conclude, therefore, is what share of the bias is explained by all of these elements together. The regression includes controls for average education level, average durable good ownership, and average ideal sex ratio reported (all three are measured at the sex-region level). The data here are limited to 1992. The explanatory power is similar to what would be expected based on the other analyses. About 40% of the sex imbalance is explained by these components.

There are obvious limitations to this approach. Nevertheless, the results are roughly consistent with the previous ones. At least some significant share of the sex imbalance—perhaps close to one-half—seems to be explained by two factors: vaccination and food intake. Of course, this result implies that at least half of the imbalance remains unexplained. One possibility is that with more accurate data, more of the imbalance could be explained. Another possibility is that there are important elements not considered here—for example, direct parental intervention.

## CONCLUSION

In this article, I use a new methodology to analyze the proximate sources of excess female mortality in India. I argue that childhood is the most crucial time period in the sense that by age 5, there is enough excess female mortality to explain the entire sex imbalance in India's population. During this period, sex differences in both vaccination and nutrition play

10. This analysis is run using India only. The sample sizes for Africa are much smaller, allowing for only a very limited number of regions, making the comparison difficult. For simplicity, I assume that the coefficient on *girl* in the regression should be zero, understanding that this is not exactly correct.

**Table 6. Regional Analysis of Proximate Causes (dependent variable is share of girls / boys in the region who died at ages 1–5)**

Explanatory Variables	Without Controls		With Controls for Vaccination or Malnourishment	
	Coefficient	<i>t</i> Statistic	Coefficient	<i>t</i> Statistic
Girl	0.0156**	4.03	0.0090*	2.50
Vaccination			0.0001	0.16
% Severely Malnourished			0.0289	0.94
Vaccinations × Malnourishment			0.0027	0.03
Average Mother's Education	−0.0015	−1.22	−0.0018†	−1.67
Average Wealth (durables)	−0.0029	−0.68	−0.0031	−0.81
Average Ideal Sex Ratio	0.0411**	3.73	0.0111	1.05
Constant	−0.0297	−1.60	0.0179	1.02
Number of Observations	339		338	
Share Explained	41.7			

*Notes:* Each observation represents a sex region, and the dependent variable is the share of children born 5–10 years ago in that sex-region group who died between the ages of 1 and 5 years (conditional on reaching age 1). The measure of vaccination represents the first principal component of indicators for having gotten seven vaccines: measles, DPT (1, 2, 3), polio (1, 2), and BCG. The data are from the 1992–1993 NFHS only. Ideal sex ratio is the average ideal number of boys divided by the average ideal number of girls.

†  $p < .10$ ; \*  $p < .05$ ; \*\*  $p < .01$

a large role in the excess mortality. I find that roughly 50% of the sex imbalance remains unexplained by differences in vaccination, nutrition, or medical care.

The first of these results stands somewhat in contrast to the situation in China, where high sex ratios at birth or immediately after seem to drive high sex ratios in the population as a whole. Das Gupta et al. (2003) also noted the apparent differences in these situations: the higher sex ratio in China seems to be driven by sex-selective abortion or sex-selective infanticide, but the situation in India (as demonstrated here) points more to childhood neglect. Understanding why these patterns differ is beyond the scope of this article. However, it may be a fruitful direction for future work, especially because it suggests that the growth of sex-selective technologies may have different effects in different areas.

The second result here—the importance of vaccinations, in particular—may shed some light on the patterns seen in the first result. The age breakdown shows that although mortality is quite high from ages 0 to 6 months, this time period does not seem to play a role in explaining excess female mortality. Initially, this contrast between the role of this time period in levels versus its role in differences may seem puzzling, but the role of vaccinations is consistent with this. Because vaccinations do not take place until a few months after birth (most between 6 and 12 months), one would expect differences attributable to vaccination to appear later in childhood, as shown in the data.

The results here have clear, and potentially important, policy implications. There has been significant focus in India on changing preferences—for example, encouraging people to put greater value on women, promoting female schooling, and so on. These are clearly useful goals. However, in the shorter run (before individual preferences can be changed), the results here argue that in particular, investments in vaccination for girls would have a direct effect on excess female mortality. Given a choice between the child investments discussed here, focus on universal (i.e., not sex-biased) vaccination would have the largest effect on mortality.

It is worth noting that in recent years, the availability of sex-selective abortion has shifted some sex selection to the period before birth (Jha et al. 2006). If pre-birth sex selection is less costly than neglect, then one would expect, in the limit, no post-birth treatment differences by sex. If parents can choose the sex of their child with certainty, then every girl who is born would be wanted—as would every boy—and one would not expect neglect after birth. It remains to be seen, however, just how large a phenomenon prenatal sex selection will become. The fact that gender differences in mortality show up not immediately at birth (i.e., from infanticide) but later in childhood (from neglect) may limit the eventual role for sex-selective abortion, which is a close substitute for infanticide, but not the role for neglect. Families may not have strong enough preferences to move to sex-selective abortion, but may still engage in less immediately obvious forms of discrimination, such as lack of vaccination. If this is true, then even where sex-selective abortion is available, the policy issues outlined here will be salient. Further, regardless of the long-run situation, it is clear that in the short run, investments in health care for girls could save thousands of lives in India.

## REFERENCES

- Aaby, P., H. Jensen, M.-L. Garly, C. Balé, C. Martins, and I. Lisse. 2002. "Vaccinations and Child Survival in a War Situation With High Mortality: Effect of Gender." *Vaccine* 21(1–2):15–20.
- Agnihotri, S. 2000. *Sex Ratio Patterns in the Indian Population: A Fresh Exploration*. New Delhi, India: Sage.
- Agnihotri, S., R. Palmer-Jones, and A. Parikh. 2002. "Missing Women in Indian Districts: A Quantitative Analysis." *Structural Change and Economic Dynamics* 13:285–314.
- Basu, A. 1989. "Is Discrimination in Food Really Necessary for Explaining Sex Differentials in Childhood Mortality?" *Population Studies* 43:193–210.
- Borooah, V. 2004. "Gender Bias Among Children in India in Their Diet and Immunization Against Disease." *Social Science and Medicine* 58:1719–31.
- Chen, L., A. Chowdhury, and S. Huffman. 1980. "Anthropometric Assessment of Energy-Protein Malnutrition and Subsequent Risk of Mortality Among Pre-School Aged Children." *American Journal of Clinical Nutrition* 33:1836–45.
- Coale, A. 1991. "Excess Female Mortality and the Balance of the Sexes: An Estimate of the Number of 'Missing Females.'" *Population and Development Review* 17:517–23.
- Coale, A., P. Demeny, and B. Vaughan. 1983. *Regional Model Life Tables and Stable Populations*. Princeton, NJ: Princeton University Press.
- Das Gupta, M. 1987. "Selective Discrimination Against Female Children in Rural Punjab, India." *Population and Development Review* 13:77–100.
- Das Gupta, M. and M. Bhat. 1997. "Fertility Decline and Increased Manifestation of Sex Bias in India." *Population Studies* 51:307–15.
- Das Gupta, M., Z. Jiang, B. Li, and W. Chung. 2003. "Why Is Son Preference So Persistent in East and South Asia? A Cross-Country Study of China, India and the Republic of Korea." *Journal of Development Studies* 40:153–87.
- Dyson, T. 1984. "Excess Male Mortality in India." *Economic and Political Weekly* 19:422–26.
- Griffiths, P., Z. Matthews, and A. Hinde. 2002. "Gender, Family, and the Nutritional Status of Children in Three Culturally Contrasting States of India." *Social Science and Medicine* 55:775–90.
- Jha, P., R. Kumar, P. Vasa, N. Dhingra, D. Thiruchelvam, and R. Moineddin. 2006. "Low Male-to-Female Sex Ratio of Children Born in India: National Survey of 1.1 Million Households." *Lancet* 367:211–18.
- Kishor, S. 1993. "'May God Give Sons to All': Gender and Child Mortality in India." *American Sociological Review* 58:247–65.
- Klasen, S. 1994. "Missing Women' Reconsidered." *World Development* 22:1061–71.
- Klasen, S. and C. Wink. 2002. "A Turning Point in Gender Bias in Mortality? An Update on the Number of Missing Women." *Population and Development Review* 28:285–312.



- Mishra, V., T.K. Roy, and R. Retherford. 2004. "Sex Differentials in Childhood Feeding, Health Care, and Nutritional Status in India." *Population and Development Review* 30:269–95.
- Murthi, A.-C., G. Mamta, and J. Dreze. 1995. "Mortality, Fertility and Gender Bias in India: A District Level Analysis." *Population and Development Review* 4:745–82.
- Padmanabha, P. 1982. "Mortality in India: A Note on Trends and Implications." *Economic and Political Weekly* 17:1285–90.
- Pande, R. 2003. "Selective Gender Differences in Childhood Nutrition and Immunization in Rural India: The Role of Siblings." *Demography* 40:395–418.
- Preston, S. and M. Bhat. 1984. "New Evidence on Fertility and Mortality Trends in India." *Population and Development Review* 10:481–503.
- Qian, N. 2008. "Missing Women and the Price of Tea in China: The Effect of Sex-Specific Earnings on Sex Imbalance." *Quarterly Journal of Economics* 123:1251–85.
- Rosenzweig, M. and T.P. Schultz. 1982. "Market Opportunities, Genetic Endowments and Intrafamily Resource Distribution: Child Survival in India." *American Economic Review* 72:803–15.
- Sen, A. 1990. "More Than 100 Million Women Are Missing." *New York Review of Books*, December 20.
- . 1992. "Missing Women." *British Medical Journal* 304:587–88.
- Visaria, P. 1971. "The Sex Ratio of the Population of India." *Monograph No. 10, Census of India*. Office of the Registrar General, New Delhi, India.
- Wells, J. 2000. "Natural Selection and Sex Differences in Morbidity and Mortality in Early Life." *Journal of Theoretical Biology* 202:65–76.